Simulating all non-signalling correlations via classical or quantum theory with negative probabilities

Sabri W. Al-Safi*
DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

Anthony J. Short†
H. H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK

Many-party correlations between measurement outcomes in general probabilistic theories are given by conditional probability distributions obeying the non-signalling condition. We show that any such distribution can be obtained from classical or quantum theory, by relaxing positivity constraints on either the mixed state shared by the parties, or the local functions which generate measurement outcomes. Our results apply to generic non-signalling correlations, but in particular they yield two distinct quasi-classical models for quantum correlations.

I. INTRODUCTION

Quantum theory predicts many strange phenomena, but non-local correlations are perhaps the most intriguing. On one hand, quantum theory is non-signalling: local measurements made by separate observers on a joint quantum state cannot convey information from one observer to another. However, the outcome statistics of these quantum measurements can correlate in such a way as to defy any classical explanation based on local influences [1].

Stronger non-locality often leads to greater aptitude in information-theoretic tasks [2–5], and consequently much work has been done to understand the non-local power of quantum theory [6–10]. Furthermore, the discovery of non-local correlations which are not achievable in quantum theory, yet still non-signalling [11], has provoked a more general study of non-local correlations and the general probabilistic theories which generate them [12–14].

In this paper, we will explore various models which are capable of generating any non-signalling correlation, yet which closely mirror the structure of classical and quantum theory. The key modification in each case is to drop a positivity requirement on one of the mathematical objects in the theory; this corresponds to allowing outcome probabilities to be negative for some unperformed measurements. Similar models have arisen elsewhere in the study of quantum mechanics - for example in the Wigner phase space representation of quantum states [15], in analysis of the EPR paradox [16], and in the context of quantum information theory [17] - and there have also been attempts to provide a physical or mathematical interpretation to the notion of negative probabilities [18, 19]. Our analysis extends to correlations arising from any non-signalling theory, and may likewise be of benefit in the study of these theories.

In the context of local models based on classical probability theory, we will show that any non-signalling correlation can be generated if either the joint probability distribution over states, or the local conditional probability distributions associated with measurements, are allowed to be negative. In the former case, this is essentially an alternative and constructive proof of the result that non-signalling states of a joint system lie in the affine hull of the pure product states [12, 20]. In the latter case, the result is more surprising, because only the positivity of the *local* measurements is modified, yet arbitrary *non-local* correlations can be generated.

In the context of quantum theory, Acín *et al* [21] have shown that the standard Born trace rule can be extended to generate any non-signalling correlation, by allowing the quantum state (usually represented by a positive density operator) to be a non-positive operator. We employ our classical framework to give an alternative proof, and a slight extension, of this result. We also derive a dual result that any non-signalling correlation can be generated if the local quantum measurement operators are allowed to be non-positive.

Acín *et al*'s result has recently been used in proving that any physical theory for which the local structure is identical to that of qubits, and which admits at least one continuous, reversible interaction, must have the global structure specified by quantum theory [22]. Likewise, the ability to represent the correlations of a broad class of theories in a similar way to quantum or classical correlations may provide a powerful tool in analysing such theories, and identifying the unique properties of quantum theory.

*Electronic address: S.W.Al-Safi@damtp.cam.ac.uk
†Electronic address: tony.short@bristol.ac.uk

II. SETUP

In order to study experimental correlations, an abstract framework is commonly used in which little is assumed about the underlying physics. Consider an experiment involving N systems, in which an observer at system k chooses one of a finite set of possible measurements, indexed by $x_k \in \{1, 2, \dots, X_k\}$, and records one of a finite set of possible outcomes, indexed by $a_k \in \{1, 2, \dots, A_k\}$. We will denote the set of possible values for k by $[N] \equiv \{1, 2, \dots, N\}$. The experiment may be characterized by the conditional probability distribution $p(a_1, \dots, a_N|x_1, \dots, x_N)$ on the measurement outcomes, given the inputs.

The correlations obtained from such an experiment are *local* if they can be generated by a local classical model, where each system has its own state which individually determines the probability of measurement outcomes on it, and where these states are distributed according to some joint probability distribution. Specifically, $p(a_1, \dots, a_N|x_1, \dots, x_N)$ is local if there exists a joint probability distribution $p_\Lambda(\lambda_1, \dots, \lambda_N)$ for the local state $\lambda_k \in \Lambda_k$ of each system k (where each Λ_k is a finite set), and a conditional probability distribution $p_k(a_k|x_k, \lambda_k)$ for each k, such that

$$p(a_1, \dots, a_N|x_1, \dots, x_N) = \sum_{\lambda_1,\dots,\lambda_N} p_1(a_1|x_1, \lambda_1) \cdots p_N(a_N|x_N, \lambda_N) p_\Lambda(\lambda_1, \dots, \lambda_N) \tag{1}$$

This definition of locality is equivalent to the more standard requirement that $p(a_1, \dots, a_N|x_1, \dots, x_N)$ lies in the convex hull of the product conditional probability distributions $p_1(a_1|x_1) \cdots p_N(a_N|x_N)$ [14], or that there is a shared state λ, obtained with probability $p(\lambda)$, which deterministically specifies the outcomes of all local measurements [1].

The discomfiting truth is that not all experiments on quantum states generate local correlations [1]. Instead, the results of quantum experiments are given in general by the formula

$$p(a_1, \dots, a_N|x_1, \dots, x_N) = \mathrm{tr}\left(\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\right)\rho\right), \tag{2}$$

where ρ is a density matrix (satisfying $\rho \geq 0$ and $\mathrm{tr}(\rho) = 1$), and $M^{(k)}_{a_k|x_k}$ are measurement operators comprising some positive-operator valued measure (POVM) for each k (satisfying $M^{(k)}_{a_k|x_k} \geq 0$ and $\sum_{a_k} M^{(k)}_{a_k|x_k} = I^{(k)}$, the identity on system k).

In section III we show that all non-signalling correlations can be represented in the form of (1), as long as either (i) the classical state p_Λ, or (ii) the classical measurement operators p_k are allowed to contain negative components. We then show in section IV that all non-signalling correlations can be represented in the form of (2), as long as either (i) the quantum state ρ, or (ii) the quantum measurement operators $M^{(k)}_{a_k|x_k}$, are allowed to be non-positive operators.

III. CLASSICAL RESULTS

In order to analyse the classical case, it will be helpful to define an analogue of a probability distribution, in which the entries are allowed to be negative; we will refer to this as a quasiprobability distribution. A function $\tilde{p} : C_1 \times \cdots \times C_k \to \mathbb{R}$, where $|C_i| < \infty$ for all i, is a *joint quasiprobability distribution* if and only if it obeys the normalisation condition

$$\sum_{c_1,\dots,c_N} \tilde{p}(c_1, \dots, c_N) = 1 \tag{3}$$

(We will only consider joint distributions, and will therefore drop the use of the word 'joint'). Similarly, a conditional quasiprobability distribution is a real function $\tilde{p}(c_1, \dots, c_K|z_1, \dots, z_L)$ which is a quasiprobability distribution for each fixed choice of $z_1, \dots, z_L$. Note that any (conditional) quasiprobability distribution which is non-negative for all values of its arguments is also a (conditional) probability distribution. For clarity we will use tildes throughout to represent quasiprobability distributions.

We wil say that a conditional quasiprobability distribution $\tilde{p}(a_1, \dots, a_N|x_1, \dots, x_N)$ is *non-signalling* if, for any $k \in [N]$ and choice of $x_1, \dots, x_N$, the sum $\sum_{a_k} \tilde{p}(a_1, \dots, a_N|x_1, \dots, x_N)$ is independent of the value of x_k. Non-signalling distributions have well-defined reduced distributions: for any non-empty subset $\mathcal{S} = \{i_1, \dots, i_M\} \subseteq [N]$ let $\mathbf{a}_\mathcal{S}$ and $\mathbf{x}_\mathcal{S}$ denote the reduced strings $(a_{i_1}, \dots, a_{i_M})$, $(x_{i_1}, \dots, x_{i_M})$; the *marginal distribution* may then be defined as

$$\tilde{p}(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S}) = \sum_{a_i \,:\, i \notin \mathcal{S}} \tilde{p}(a_1, \dots, a_N|x_1, \dots, x_N) \tag{4}$$

for some arbitrary choice of x_i for $i \notin \mathcal{S}$.

Lemma 1. *A non-signalling, conditional quasiprobability distribution $\tilde{p}(a_1,\ldots,a_N|x_1,\ldots,x_N)$ is uniquely characterized by the complete set of marginal distributions $\tilde{p}(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S})$ for which $a_{i_k} < A_{i_k}\,\forall i_k \in \mathcal{S}$, where $\mathcal{S}$ ranges over all subsets of $[N]$.*

Proof. This lemma is an explicit statement in terms of quasiprobabilities of a result used in [21]. Suppose that $\tilde{p}(a_1,\ldots,a_N|x_1,\ldots,x_N)$ and $\tilde{p}'(a_1,\ldots,a_N|x_1,\ldots,x_N)$ are non-signalling conditional quasiprobability distributions, whose marginals $\tilde{p}(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S})$ and $\tilde{p}'(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S})$ agree whenever $a_{i_k} < A_{i_k}\,\forall i_k \in \mathcal{S}$, for all $\mathcal{S}$. We argue that $\tilde{p}(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S})$ and $\tilde{p}'(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S})$ agree for all choices of a_i, by induction on $n = \#\{i_k \in \mathcal{S} : a_{i_k} = A_{i_k}\}$. The case $n=0$ holds trivially; for $n>0$, without loss of generality, suppose that $a_{i_1} = A_{i_1}$, then,

$$\begin{aligned}
\tilde{p}(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S}) &= \tilde{p}(A_{i_1},\ldots,a_{i_M}|x_{i_1},\ldots,x_{i_M}) && (5)\\
&= \tilde{p}(a_{i_2},\ldots,a_{i_M}|x_{i_2},\ldots,x_{i_M}) - \sum_{a_{i_1}<A_{i_1}} \tilde{p}(a_{i_1},\ldots,a_{i_M}|x_{i_1},\ldots,x_{i_M}) && (6)\\
&= \tilde{p}'(a_{i_2},\ldots,a_{i_M}|x_{i_2},\ldots,x_{i_M}) - \sum_{a_{i_1}<A_{i_1}} \tilde{p}'(a_{i_1},\ldots,a_{i_M}|x_{i_1},\ldots,x_{i_M}) && (7)\\
&= \tilde{p}'(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S}). && (8)
\end{aligned}$$

Line (6) follows from the definition of the marginal distribution, and line (7) follows from applying the induction hypothesis to the $n-1$ case. Setting $\mathcal{S} = [N]$ then proves the lemma. ◻

This lemma tells us that as long as the non-signalling property is obeyed, we can restrict our attention to a subset of all possible measurement outcomes: this is very helpful in proving the following Theorems.

Theorem 1. *(Non-positive classical measurements) An N-partite conditional probability distribution $p(a_1,\ldots,a_N|x_1,\ldots,x_N)$ is non-signalling if and only if it can be represented in the form*

$$p(a_1,\ldots,a_N|x_1,\ldots,x_N) = \sum_{\lambda_1,\ldots,\lambda_N} \tilde{p}_1(a_1|x_1,\lambda_1)\cdots\tilde{p}_N(a_N|x_N,\lambda_N)p_\Lambda(\lambda_1,\ldots,\lambda_N) \tag{9}$$

where $p_\Lambda(\lambda_1,\ldots,\lambda_N)$ is a probability distribution, and $\tilde{p}_k(a_k|x_k,\lambda_k)$ is a conditional quasiprobability distribution for each k.

Proof. Firstly, note that as the $\tilde{p}_k(a_k|x_k,\lambda_k)$ are conditional quasiprobability distributions, they must satisfy $\sum_{a_k}\tilde{p}_k(a_k|x_k,\lambda_k) = 1$ (independent of x_k). It follows that the distribution given by (9) is non-signalling. Conversely, to prove that all non-signalling distributions $\tilde{p}$ can be written in the form of (9), we take each Λ_k to be the set of ordered pairs $[a'_k, x'_k]$ consisting of the allowed measurement choices and outputs for system k. We then set

$$\begin{aligned}
p_\Lambda([a'_1,x'_1],\ldots,[a'_N,x'_N]) &= \frac{p(a'_1,\ldots,a'_N|x'_1,\ldots,x'_N)}{X_1X_2\cdots X_N} && (10)\\
\tilde{p}_k(a_k|x_k,\lambda_k) &= \begin{cases} X_k\delta_{\lambda_k,[a_k,x_k]} & \text{if} \quad a_k < A_k\\ 1-\sum_{a<A_k} X_k\delta_{\lambda_k,[a,x_k]} & \text{if} \quad a_k = A_k \end{cases} && (11)
\end{aligned}$$

With these assignments, $p_\Lambda(\lambda_1,\ldots,\lambda_N)$ is a probability distribution and $\tilde{p}_k(a_k|x_k,\lambda_k)$ is a conditional quasiprobability distribution. However, $\tilde{p}_k(A_k|x_k,[1,x_k]) = 1 - X_k < 0$ whenever $X_k > 1$, hence $\tilde{p}_k(a_k|x_k,\lambda_k)$ will not usually be a valid conditional probability distribution.

It remains to show that the given values for $p_\Lambda(\lambda_1,\ldots,\lambda_N)$ and $\tilde{p}_k(a_k|x_k,\lambda_k)$ satisfy (9). Consider the quasiprobability distribution p' given by

$$p'(a_1,\ldots,a_N|x_1,\ldots,x_N) = \sum_{\lambda_1,\ldots,\lambda_N} \tilde{p}_1(a_1|x_1,\lambda_1)\cdots\tilde{p}_N(a_N|x_N,\lambda_N)p_\Lambda(\lambda_1,\ldots,\lambda_N). \tag{12}$$

It is straightforward to check that p' is non-signalling and that $p(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S}) = p'(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S})$ for all subsets $\mathcal{S} \subseteq [N]$, and strings $\mathbf{a}_\mathcal{S}$ and $\mathbf{x}_\mathcal{S}$ with $a_i < A_i\ \forall i \in \mathcal{S}$. For example, when $\mathcal{S} = \{1,2\}$ we have

$$\begin{aligned}
p'(a_1,a_2|x_1,x_2) &= \sum_{\lambda_1,\ldots,\lambda_N} X_1\delta_{\lambda_1,[a_1,x_1]}X_2\delta_{\lambda_2,[a_2,x_2]}p_\Lambda(\lambda_1,\ldots,\lambda_N).\\
&= X_1X_2\frac{p(a_1,a_2|x_1,x_2)}{X_1X_2} = p(a_1,a_2|x_1,x_2). && (13)
\end{aligned}$$

By lemma 1, this is enough to conclude that p' is indeed the probability distribution p, hence (9) holds. ◻

Theorem 2. *(Non-positive classical states) An N-partite conditional probability distribution $p(a_1,\ldots,a_N|x_1,\ldots,x_N)$ is non-signalling if and only if it can be represented in the form*

$$p(a_1,\ldots,a_N|x_1,\ldots,x_N) = \sum_{\lambda_1,\ldots,\lambda_N} p_1(a_1|x_1,\lambda_1)\cdots p_N(a_N|x_N,\lambda_N)\tilde{p}_\Lambda(\lambda_1,\ldots,\lambda_N) \tag{14}$$

where $\tilde{p}_\Lambda(\lambda_1,\ldots,\lambda_N)$ is a quasiprobability distribution, and $p_k(a_k|x_k,\lambda_k)$ is a conditional probability distribution for each k.

Proof. As each $p_k(a_k|x_k,\lambda_k)$ is a conditional probability distribution, it is clear that summing over a_k removes any dependence on x_k on the right-hand side, hence the distribution $p(a_1,\ldots,a_N|x_1,\ldots,x_N)$ is non-signalling. It remains to be shown that we can represent any non-signalling distribution in the form of (14). To achieve this, we take each Λ_k to be the set of ordered pairs $[a'_k,x'_k]$ as before, along with an additional value which we will refer to as ξ_k, so that there are $A_k\cdot X_k+1$ possible choices of λ_k for each k. For a string $(\lambda_1,\ldots,\lambda_N)\in\Lambda_1\times\cdots\times\Lambda_N$, let $\mathcal{S}=\{i\in[N]\,:\,\lambda_i\neq\xi_i\}$ (i.e. the set of indices i for which $\lambda_i\neq\xi_i$) and define

$$\tilde{p}_\Lambda(\lambda_1,\ldots,\lambda_N) = \left[\Pi_{i\notin\mathcal{S}}(1-X_i)\right]p(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S}) \tag{15}$$

so that, for example,

$$\begin{aligned}
\tilde{p}_\Lambda([a'_1,x'_1],[a'_2,x'_2],\ldots,[a'_N,x'_N]) &= p(a'_1,\ldots,a'_N|x'_1,\ldots,x'_N) && (16)\\
\tilde{p}_\Lambda(\xi_1,[a'_2,x'_2],\ldots,[a'_N,x'_N]) &= (1-X_1)p(a'_2,\ldots,a'_N|x'_2,\ldots,x'_N) && (17)\\
\vdots &= \vdots && (18)\\
\tilde{p}_\Lambda(\xi_1,\xi_2,\ldots,\xi_N) &= (1-X_1)(1-X_2)\cdots(1-X_N). && (19)
\end{aligned}$$

To show that this is a quasiprobability distribution, note that

$$\begin{aligned}
\sum_{\lambda_1,\ldots,\lambda_N}\tilde{p}_\Lambda(\lambda_1,\ldots,\lambda_N) &= \sum_{\mathcal{S}\subseteq[N]}\sum_{\mathbf{a}_\mathcal{S},\mathbf{x}_\mathcal{S}}\left[\Pi_{i\notin\mathcal{S}}(1-X_i)\right]p(\mathbf{a}_\mathcal{S}|\mathbf{x}_\mathcal{S}) && (20)\\
&= \sum_{\mathcal{S}\subseteq[N]}\left[\Pi_{i\notin\mathcal{S}}(1-X_i)\right]\left[\Pi_{j\in\mathcal{S}}X_j\right] && (21)\\
&= \Pi_{i\in[N]}\left(X_i+(1-X_i)\right)=1. && (22)
\end{aligned}$$

The $p_k(a_k|x_k,\lambda_k)$ are defined to be the following conditional probability distributions:

$$p_k(a_k|x_k,\lambda_k) = \begin{cases}\delta_{\lambda_k,[a_k,x_k]} & \text{if} \quad a_k<A_k\\ 1-\sum_{a<A_k}\delta_{\lambda_k,[a,x_k]} & \text{if} \quad a_k=A_k\end{cases}. \tag{23}$$

As in the previous case, it is a straightforward check that substituting (15) and (23) into (12) gives a non-signalling conditional quasiprobability distribution whose marginals agree with all the marginals of the distribution $p(a_1,\ldots,a_N|x_1,\ldots,x_N)$ in the case $a_i<A_i\ \forall i$. Hence, again by lemma 1 the initial probability distribution is recovered. Note that the measurement probabilites $p_k(a_k|x_k,\lambda_k)$ always equal 0 or 1 in this case, hence the measurements are in fact deterministic. □

IV. QUANTUM RESULTS

By a careful construction of states and measurements, our classical results immediately imply quantum corollaries. In particular, we can use Theorem 1 to prove

Corollary 1. *(Non-positive quantum measurements) An N-partite conditional probability distribution $p(a_1,\ldots,a_N|x_1,\ldots,x_N)$ is non-signalling if and only if it can be represented in the form*

$$p(a_1,\ldots,a_N|x_1,\ldots,x_N) = \mathrm{tr}\left(\left(\tilde{M}^{(1)}_{a_1|x_1}\otimes\cdots\otimes\tilde{M}^{(N)}_{a_N|x_N}\right)\rho\right), \tag{24}$$

where $\tilde{M}^{(k)}_{a_k|x_k}$ are Hermitian operators satisfying $\sum_{a_k}\tilde{M}^{(k)}_{a_k|x_k}=I$ for each k, and ρ is a density operator (satisfying $\rho\geq 0,\mathrm{tr}(\rho)=1$). Furthermore, this representation can be chosen such that the operators $\tilde{M}^{(1)}_{a_1|x_1}\otimes\cdots\otimes\tilde{M}^{(N)}_{a_N|x_N}$ and ρ all commute.

Proof. Summing the right-hand side of (24) over a_k, it is clear that any p which can be represented in this way is non-signalling. To prove the converse, we use the results of Theorem 1. To each system k, we assign a Hilbert space spanned by the orthonormal basis $\{|\lambda_k\rangle | \lambda_k \in \Lambda_k\}$, where Λ_k is the set of all ordered pairs $[a_k, x_k]$. We then take

$$\rho = \sum_{\lambda_1,\ldots,\lambda_N} p_\Lambda(\lambda_1,\ldots,\lambda_N)\, |\lambda_1\rangle\langle\lambda_1| \otimes |\lambda_2\rangle\langle\lambda_2| \otimes \cdots \otimes |\lambda_N\rangle\langle\lambda_N|, \tag{25}$$

$$\tilde{M}^{(k)}_{a_k|x_k} = \sum_{\lambda_k} \tilde{p}_k(a_k|x_k,\lambda_k)\, |\lambda_k\rangle\langle\lambda_k|, \tag{26}$$

where $p_\Lambda(\lambda_1,\ldots,\lambda_N)$ and $\tilde{p}_k(a_k|x_k,\lambda_k)$ are given by (10) and (11) respectively. Substituting this choice of operators into (24) leads to exactly the same equation as (9): therefore the result follows directly from Theorem 1. □

Similarly, from Theorem 2 we may obtain a re-derivation of Acín *et al*'s result [21], with the slight strengthening that all the operators may be chosen to commute.

Corollary 2. *(Non-positive quantum states) An N-partite conditional probability distribution* $p(a_1,\ldots,a_N|x_1,\ldots,x_N)$ *is non-signalling if and only if it can be represented in the form*

$$p(a_1,\ldots,a_N|x_1,\ldots,x_N) = \mathrm{tr}\left(\left(M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}\right)\tilde{\rho}\right), \tag{27}$$

where $M^{(k)}_{a_k|x_k}$ *are POVM elements (positive operators satisfying* $\sum_{a_k} M^{(k)}_{a_k|x_k} = I$*) for each* k*, and* $\tilde{\rho}$ *is a Hemitian operator satisfying* $\mathrm{tr}(\rho) = 1$*. Furthermore this representation can be chosen such that the operators* $M^{(1)}_{a_1|x_1} \otimes \cdots \otimes M^{(N)}_{a_N|x_N}$ *and* $\tilde{\rho}$ *all commute.*

Proof. Again we see that any such p is non-signalling by summing the right-hand side of (27) over a_k. To prove the converse, we assign a Hilbert space spanned by the orhonormal basis $\{|\lambda_k\rangle | \lambda_k \in \Lambda_k\}$ to each system k, where Λ_k is the set of all ordered pairs $[a_k, x_k]$ and the extra element ξ_k (as in the proof of Theorem 2). We then take

$$\tilde{\rho} = \sum_{\lambda_1,\ldots,\lambda_N} \tilde{p}_\Lambda(\lambda_1,\ldots,\lambda_N)\, |\lambda_1\rangle\langle\lambda_1| \otimes |\lambda_2\rangle\langle\lambda_2| \otimes \cdots \otimes |\lambda_N\rangle\langle\lambda_N|, \tag{28}$$

$$M^{(k)}_{a_k|x_k} = \sum_{\lambda_k} p_k(a_k|x_k,\lambda_k)\, |\lambda_k\rangle\langle\lambda_k|, \tag{29}$$

where $\tilde{p}_\Lambda(\lambda_1,\ldots,\lambda_N)$ and $p_k(a_k|x_k,\lambda_k)$ are given by (15) and (23) respectively. Substituting these into (27) and applying Theorem 2 proves the corollary. □

In both these theorems, the set of quantum correlations can be recovered by adding the requirement that the operators $\tilde{M}^{(k)}_{a_k|x_k}$ or $\tilde{\rho}$ are positive respectively.

V. DISCUSSION

One way of viewing the mixture of states p_Λ of a local distribution (1) is as a result of one's ignorance of the particular value of $(\lambda_1,\ldots,\lambda_N)$. From this perspective, the results of Theorem 1 are particularly surprising: in this case the state $p_\Lambda(\lambda_1,\ldots,\lambda_N)$ is a standard mixture of product states (which one would normally think of as 'local'), and all the measurement distributions $\tilde{p}_k(a_k|x_k,\lambda_k)$ are local objects. One might therefore wonder how this can generate non-local correlations at all. The explanation is that the measurements do not yield positive outcome probabilities for each component of the mixed state. Hence we can no longer think of the state as an ignorance mixture of allowed local states, but rather as a non-local object itself.

As it stands, there is no discernible difference in using our procedures to generate a local correlation, compared with using them to generate a non-local correlation. In all cases the resulting distributions or operators will contain negative components, as long as at least one party has more than one measurement choice. It would be interesting to see whether there exists a procedure similar to ours which outputs genuine probability distributions when generating local correlations. The question also arises whether quantum correlations might have some special status when represented in the form of classical quasiprobability representations as in (9) and (14): perhaps, for example, the negative values in the distributions can be bounded if the correlation is quantum. However, this seems likely to be difficult, given the apparent difficulty of finding simple conditions for the quantum realizability of correlations[23].

It is interesting to examine the efficiency of our representations. Note that the number of hidden variables we use is $\Pi_{k=1}^{N} A_k X_k$ in Theorem 1 and $\Pi_{k=1}^{N}(A_k X_k + 1)$ in Theorem 2. This can be reduced by noting that whenever λ_k is equal to either ξ_k or $[A_k, x_k]$, each local conditional probability distribution takes the same form: $p(a_k|x_k, \lambda_k) = \delta_{a_k, A_k}$. Hence this specific set of states can be combined into a single state η_k. To preserve normalization, whenever $\lambda_k = \eta_k$ the probabilities (10) and (15) must be summed over all the combined states. We do not use this compression of the state space in the main presentation of the theorems for clarity. However, this would reduce the total number of states in both classical theorems to $\Pi_{k=1}^{N}((A_k - 1)X_k + 1)$, which is the same as the number of real parameters used to specify a non-signalling probability distribution via its marginals in Lemma 1 (in practice, one less parameter is needed due to normalisation). In the quantum case, our representations use a quadratically larger Hilbert space than that of [21], with the payoff that all the operators commute.

We have shown that by taking either classical probability theory or quantum theory as a starting point, and by relaxing a positivity constraint on *either* the statistics of measurement outcomes, *or* the probabilistic mixing of states, one generates all non-signalling correlations. It would be interesting to investigate whether other general probabilistic theories [14] can be modified in a similar manner to yield all non-signalling correlations, or whether this is particular to theories which, like quantum theory, contain classical theory as a special case.

Acknowledgments. AJS acknowledges support from the Royal Society. SWA is funded by an EPSRC grant, and thanks Paul Skrzypczyk and James Yearsley for helpful discussions.

---

[1] J. S. Bell, “On the Einstein-Podolsky-Rosen paradox,” *Physics*, vol. 1, pp. 195–200, 1965.

[2] W. van Dam, “Implausible consequences of superstrong nonlocality.” quant-ph/0501159, 2005.

[3] N. Linden, S. Popescu, A. J. Short, and A. Winter, “Quantum nonlocality and beyond: Limits from nonlocal computation,” *Phys. Rev. Lett.*, vol. 99, p. 180502, 2007.

[4] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, “Information causality as a physical principle,” *Nature*, vol. 461, pp. 1101–1104, 2009.

[5] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio, “Guess your neighbor’s input: A multipartite nonlocal game with no quantum advantage,” *Phys. Rev. Lett.*, vol. 104, p. 230404, Jun 2010.

[6] B. Tsirelson, “Quantum generalizations of Bell’s inequality,” *Lett. Math. Phys.*, vol. 4, pp. 93–100, 1980.

[7] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger “Limit on nonlocality in any world in which communication complexity is not trivial,” *Phys. Rev. Lett.*, vol. 96, p. 250401, Jun 2006.

[8] S. W. Al-Safi and A. J. Short, “Information causality from an entropic and a probabilistic perspective,” *Phys. Rev. A*, vol. 84, p. 042323, Oct 2011.

[9] O. C. Dahlsten, D. Lercher, and R. Renner, “Tsirelson’s bound from a generalized data processing inequality,” *New Journal of Physics*, vol. 14, p. 063024., Jun 2012

[10] J. Allcock, N. Brunner, M. Pawlowski, and V. Scarani, “Recovering part of the boundary between quantum and nonquantum correlations from information causality,” *Phys. Rev. A*, vol. 80, p. 040103, 2009.

[11] S. Popescu and D. Rohrlich, “Quantum nonlocality as an axiom,” *Found. Phys.*, vol. 24, pp. 379–385, Mar 1994.

[12] J. Barrett, “Information processing in generalized probabilistic theories,” *Phys. Rev. A*, vol. 75, p. 032304, 2007.

[13] A. Short and J. Barrett, “Strong nonlocality: a trade-off between states and measurements,” *New Journal of Physics*, vol. 12, p. 033034, 2010.

[14] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, “Nonlocal correlations as an information-theoretic resource,” *Phys. Rev. A*, vol. 71, p. 022101, Feb 2005.

[15] E. Wigner, “On the quantum correction for thermodynamic equilibrium,” *Phys. Rev.*, vol. 40, pp. 749–759, Jun 1932.

[16] R. Feynman, “Simulating physics with computers,” *Int. J. Theor. Phys.*, vol. 21, p. 467, 1982.

[17] C. Ferrie, “Quasi-probability representations of quantum theory with applications to quantum information science” quantu-ph/1010.2701, 2011.

[18] W. Mückenheim, G. Ludwig, C. Dewdney, P. Holland, A. Kyprianidis, J. Vigier, N. Cufaro Petroni, M. Bartlett, and E. Jaynes, “A review of extended probabilities,” *Physics Reports*, vol. 133, pp. 337–401, 1986.

[19] A. Khrennikov, *Interpretations of Probability.* Walter de Gruyter, 2009.

[20] J. Degorre, M. Kaplan, S. Laplante and J. Roland, “The communication complexity of non-signaling distributions”, *Quantum Info. Comput.* , vol. 11, p. 649–676, Jul 2011.

[21] A. Acín, R. Augusiak, D. Cavalcanti, C. Hadley, J. K. Korbicz, M. Lewenstein, L. Masanes, and M. Piani, “Unified framework for correlations in terms of local quantum observables,” *Phys. Rev. Lett.*, vol. 104, p. 140404, Apr 2010.

[22] G. de la Torre, L. Masanes, A. J. Short, and M. P. Müller, “Deriving quantum theory from its local structure and reversibility,” *Phys. Rev. Lett.*, vol. 109, p. 090403, Aug 2012.

[23] M. Navascues, S. Pironio, and A. Acin, “A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations,” *New Journal of Physics*, vol. 10, p. 073013, Jul 2008.